Chapter 1

AN INTRODUCTION TO SOCIAL NETWORK DATA ANALYTICS

Charu C. Aggarwal
IBM T. J. Watson Research Center
Hawthorne, NY 10532
charu@us.ibm.com

Abstract The advent of online social networks has been one of the most exciting events in this decade. Many popular online social networks such as Twitter, LinkedIn, and Facebook have become increasingly popular. In addition, a number of multimedia networks such as Flickr have also seen an increasing level of popularity in recent years. Many such social networks are extremely rich in content, and they typically contain a tremendous amount of content and linkage data which can be leveraged for analysis. The linkage data is essentially the graph structure of the social network and the communications between entities; whereas the content data contains the text, images and other multimedia data in the network. The richness of this network provides unprecedented opportunities for data analytics in the context of social networks. This book provides a data-centric view of online social networks; a topic which has been missing from much of the literature. This chapter provides an overview of the key topics in this field, and their coverage in this book.

Keywords: Social Networks, Data Mining

1. Introduction

This chapter will provide an introduction of the topic of social networks, and the broad organization of this book. Social networks have become very popular in recent years because of the increasing proliferation and affordability of internet enabled devices such as personal computers, mobile devices and other more recent hardware innovations such as internet tablets. This is evidenced by the burgeoning popularity of many online social networks such as Twitter,
Facebook and LinkedIn. Such social networks have lead to a tremendous explosion of network-centric data in a wide variety of scenarios. Social networks can be defined either in the context of systems such as Facebook which are explicitly designed for social interactions, or in terms of other sites such as Flickr which are designed for a different service such as content sharing, but which also allow an extensive level of social interaction.

In general, a social network is defined as a network of interactions or relationships, where the nodes consist of actors, and the edges consist of the relationships or interactions between these actors. A generalization of the idea of social networks is that of information networks, in which the nodes could comprise either actors or entities, and the edges denote the relationships between them. Clearly, the concept of social networks is not restricted to the specific case of an internet-based social network such as Facebook; the problem of social networking has been studied often in the field of sociology in terms of generic interactions between any group of actors. Such interactions may be in any conventional or non-conventional form, whether they be face-to-face interactions, telecommunication interactions, email interactions or postal mail interactions.

The conventional studies on social network analysis have generally not focussed on online interactions, and have historically preceded the advent and popularity of computers or the internet. A classic example of this is the study of Milgram [18] in the sixties (well before the invention of the internet), who hypothesized the likelihood that any pair of actors on the planet are separated by at most six degrees of separation. While such hypotheses have largely remained conjectures over the last few decades, the development of online social networks has made it possible to test such hypotheses at least in an online setting. This is also referred to as the small world phenomenon. This phenomenon was tested in the context of MSN messenger data, and it was shown in [16] that the average path length between two MSN messenger users is 6.6. This can be considered a verification of the widely known rule of “six degrees of separation” in (generic) social networks. Such examples are by no means unique; a wide variety of online data is now available which has been used to verify the truth of a host of other conjectures such as that of shrinking diameters[15] or preferential attachment. In general, the availability of massive amounts of data in an online setting has given a new impetus towards a scientific and statistically robust study of the field of social networks.

---

1The shrinking diameter conjecture hypothesizes that the diameters of social networks shrink in spite of the addition of new nodes, because of an increase in the density of the underlying edges. The preferential attachment conjecture hypothesizes that new nodes and edges in the social networks are more likely to be attached to the dense regions of the network.